CHAPTER THREE

EXPERIMENT ONE

3.00 CHAPTER OVERVIEW

This chapter is divided into a number of sections. Section 3.01 is a preamble on the issue, questions and aim of Experiment 1. Section 3.02 is the introduction to the experiment and begins with a comparison of Anderson's (1991a) and Rosch's (1973) approaches to the issue of whether people mentally represent their categories as gradient structures. The possibility of graded structure in categories other than natural types is also discussed. Section 3.03 contains two accounts of the gradience effects commonly found in data such as the production frequency of category exemplars across a subject population. Section 3.04 sets out how these two accounts might be tested, since they generate differing predictions for patterns of participants' agreement at various levels of exemplar production frequency.

Section 3.05 describes the method and procedures of Experiment 1; and section 3.06 sets out the results of the first and second analyses of the data. Section 3.07 states the experiment's main findings on category-type differences, and gradient structure. Section 3.08 discusses the results' implications for further experiments.

3.01 ISSUE AND AIM OF EXPERIMENT 1

The research literature abounds with evidence for some sort of system in people's category behaviour, such as the consistent and stable ordering of exemplars within a category-extension. For example, when people are asked to give an instance of *Bird*, a majority will consistently produce *sparrow* or *robin*, some might suggest *gull*, and few (if any) will generate *penguin*. This ordering of exemplars according to production frequencies has been found to correlate strongly with the same items' response times in membership decision tasks, and with their typicality ratings (Barsalou, 1983, 1985; Barsalou & Sewell, 1985; Hampton & Gardiner, 1983; Mervis, Catlin & Rosch, 1976). Gradience effects generally, whether they are found in typicality ratings, response times or production frequencies, are a puzzle for researchers. A number of theories have been put forward to explain them, two of which will be tested in this experiment.

Two main accounts for gradience effects have emerged. The earlier group of Roschean theories concerns the structural representation of categories, and claims that people have fuzzy perceptions of the natural structure of their physical environment. Consequently, they are said to represent the objects in their world on a graded continuum of representativeness, with some objects being better examples of the prototype concept than others. As an alternative to structural representation, theories of the formal representation of categories claim that people's categorization behaviour is deterministic. People are said to be innately programmed to "discover" ontological categories inherent in the metaphysical structure of their environment. Categories are represented as formal or abstract rules (for example, algorithms) used for membership decision or exemplar production (Anderson, 1991a).

The issue raised by the two theories and investigated in this study concerns whether the individual exemplars in a category-extension are represented as an organized gradient structure, or as a bundle of haphazard instances associated with their category to a greater or smaller degree. Anderson's theory claims that the individual exemplars which constitute a category-extension are not represented at all, only the membership rule is represented (Murphy, 1993a). The two theories differ on this issue of whether people share a common understanding of the category mentally represented. Are the gradience effects found in the empirical data, and produced by tasks such as exemplar-generation, caused by the vagaries of personal experience, or because people mentally represent the individual objects and creatures found in their world?

The general aim of this experiment is to discover whether the order of instances found in a production frequency distribution of exemplars is a statistical artifact of the data collection. Anderson's approach would say that any order is due to random effects of learning and recency of experience, whilst Rosch would claim that such order represents a gradient structure. Both theories, as briefly explained in section 3.03, have an explanation for the graded distributions of production frequency values found in category extensions.

3.02 GRADIENT STRUCTURE AND CATEGORY-TYPES

The Roschean or prototype account claims that people mentally represent the gradient structure of a category's extension. Even simple categories are viewed as being part of hierarchically structured mental representations, consisting of bundles of correlated features (Rosch, Mervis, Gray, Johnson, & Boyes-Braem , 1976); or in the case of Rosch and Mervis' (1975) family resemblance theory, lists of independent features. See Appendix C for a description of both studies. Each exemplar's typicality increases as the number of similar features shared with other exemplars also increases. Thus, each individual exemplar belonging to a category can be arranged on a continuum of internal membership structure, with its place in the sequential order determined by its degree of typicality. Because any potential member is categorized according to its degree of typicality of the category, the internal category structure is *graded,* with its exemplars representing differing degrees of membership. This gradient structure is considered to be stable and unchanging, with significant agreement between subjects on the typicality ratings of instances in the same category (Rosch, 1973; Rips, Shoben & Smith, 1973; but cf. Barsalou, 1987).

Anderson's (1991a) is a more recent view of category representation than Rosch's, and he proposed a rationalist theory of adaptive cognition, where categories are disjoint sets of instances and represented as formal rules. Since all instances are considered to be equal members of the category, they are more like a bundle of objects than systematically ordered or structured. Any gradience effects are determined by how recently and how often an object was experienced (Murphy, 1993b). So gradedness is not a product of representational structure, but how frequently the object has occurred in *the individual person's* experience. This means that, since people differ in the frequency with which they might encounter an object, they are unlikely to agree or produce exemplars in any systematic order of generation.

Both Anderson and Rosch assume that the coherence of items as a category unit is based upon the similarity of their items, and such an assumption might be valid where natural category-types are concerned. Yet gradience effects have been found in the extensions of a number of category-types other than natural types, and whose instances do not share physical similarity (Barsalou, 1985, 1987; Homa, 1984; Rosch & Mervis, 1975; Smith & Medin, 1981;

Oden, 1977; 1987). For abstract categories (for example, *Works of Art*), Hampton (1981) found gradience in the instances contained in the category-extension (*paintings, sculptures*) although these instances do not share any physically similar features at all. Armstrong, Gleitman and Gleitman (1983) found gradience in the extensions of formal categories such as *Odd Numbers* and *Squares.* Barsalou and Ross (1986) found gradience in the production frequency of exemplars of property category-types, such as *Loud things, Red things.* Where ad hoc category-types are concerned, these are spontaneously constructed in order to achieve some goal, such as *Things to take on a picnic.* Ad hoc exemplars are rarely physically similar, yet they have been found to exhibit some gradedness in their judged degree of typicality (Barsalou, 1983, 1985), though this gradience is a less stable structure than found in, for example, natural category-types (Barsalou, 1987).

Empirical evidence of gradience found in the extensions of categories belonging to abstract, formal or ad hoc category-types might be indicating underlying structure in the representation of such categories, but not necessarily one of physical appearance as in natural category-types. What is true of one category-type might be untrue of another one. For this reason, three category-types (natural superordinates, property types and ad hoc types) are examined in Experiment 1, rather than natural types only, to investigate gradience effects in category-types whose items are not physically similar. The main point upon which the three category-types most differ is in the physical similarity of their items. The members of ad hoc types need not share any similarity of appearance at all; items in a property category must share the one salient feature, not necessarily a physically similar one; and the exemplars of superordinate types seem to share a number of similar features, but not always the same ones.

3.03 TWO ACCOUNTS OF GRADIENCE FOUND IN PRODUCTION FREQUENCY DATA

Production frequency data is collected by asking people to write down exemplars in response to a category name, then noting the frequency of occurrence of each response produced to each category name. Such a list becomes a production frequency distribution for each category (Cohen, Bousfield & Whitmarsh, 1957; Shapiro & Palermo, 1970; Battig & Montague, 1969). It provides an index of "instance dominance", referring to the percentage of persons who give a particular item as an exemplar for a general category (Wilkins, 1971). It also provides an index of "subject agreement" on which items are the best exemplars. Thus, a frequency distribution for *Bird* might be as follows: *magpie* 63%; *budgie* 49%; *emu* 31%; *kookaburra* 27%; *hawk* 21%. Incidentally, the item *robin* seems to be the instance produced by the majority of participants in overseas' researchers' studies. An exemplar list such as the one given above reflects the judgment of the Australian participants who produced it.

3.03.1 Learning order of items determines their later recall

The different frequencies with which people produce exemplars of a category might be explained as a product of learning order, and category structure can vary substantially as a function of it. The order of exemplars in a category might be governed by such extraneous factors as the order in which individuals study or experience the instances (Anderson & Matessa, 1990; Dickinson, 1991). The order of exemplars produced in response to a category label, then, would simply reflect the order of occurrence in which they were learned as instances of that particular category. Retrievability of an item in a production frequency task might be influenced by how much a person needs the information, how often the item has been read in written material in the past; or how recently that particular item was seen or accessed, and whether it is still active in working memory (Anderson, 1991b). None of these reasons lead to a Roschean structural account of the frequency distribution for items in a category.

Since Anderson's (1990) rational model assumes that all exemplars are equally good examples of their category, the only explanation for the presence of internal gradience amongst them is one which involves the varying frequency of experiences across different participants. This does not mean that Anderson is ignoring people's cognitive biases, or claiming that his rational model is a prescription for how people *should* calculate probabilities during decision-making or categorization (Gregson, 1991; Evans, 1991; Snow, 1991). For example, he does claim that people use base-rates during categorization, represented by category-size in memory. However, he stipulates that a distinction be made between experienced and stated base-rates. People ignore stated, abstract information gathered by reliable or official sources, but use instead personal experience. Thus, the exemplars people are likely to produce in response to a category-label will be based upon *individual,* experienced probabilities of selection from a large or small category-size.

Such an experiential account means that there would be little, if any, participant agreement about what order the exemplars should be retrieved in. After all, different people have different experiences of different objects, and certainly people differ in how often they might experience the same object. As a result, the gradience in a production frequency distribution based on data from a participant population might be a statistical artifact. Out of one hundred participants, ten people might mention *robin* first in their individual list of instances of *Bird,* twenty-five might mention *robin* third on their list, and twenty-five might mention it fifth. *Magpie* on the other hand might be consistently generated first in the list by fifty people, but not mentioned by any other participants further down the list. On this reasoning, *robin* would be higher in item dominance (that is, more frequent in occurrence) on the list of exemplars than *magpie,* even though the latter was more consistently produced first. It is possible that the norms gained from production frequency measures might simply reflect the percentage of participants who generate a certain instance to a particular category. The alternative explanation for gradience effects described in the next section assumes that production frequency norms reflect something more than a statistical artifact.

3.03.2 Typicality structures determine production order

The semantic relatedness account assumes that gradience effects reflect the degree of meaningful relationship between the specific item and its general category. The main point is that semantic relatedness causes exemplars to vary *in a predictable way* in a number of tasks, and across large subject populations. When the item is one of high typicality or strong associative strength, then participant performance in certain tasks is always enhanced. Those categories, instances or their properties which take longer for retrieval, are said to do so because they are further removed from more typical exemplars in the semantic organization of memory. They are less likely to be retrieved, and are more prone to be mistaken in retrieval than are those which are closer to one another in a category representation. Such evidence implies that a strictly organized category representation of individual exemplars exists even at the basic level of that category (Rips, Shoben & Smith 1973).

More specifically, variations in the frequency with which people generate members of categories has been found to correlate strongly with typicality. Exemplars judged as being more typical of their category are also produced more often (than atypical exemplars); or exemplars more strongly associated with the category name (as measured by speed of response) are produced more often (Barsalou, 1983, 1985; Barsalou & Sewell, 1985; Mervis, Catlin & Rosch, 1976). Because production frequency gradiences have been found to correlate so strongly with the gradiences in typicality and/or strength of association between words and their categories, the implication is that production frequency might be a measure of how instances are represented semantically in memory.

Gradience has also been found in the time taken to classify something as a category member, with typical exemplars being identified faster than atypical exemplars (Rosch, 1973; Rips, Shoben & Smith, 1973; McCloskey & Glucksberg, 1979; Smith, 1978). Also, participants take less time to verify a category statement like *a chair is a piece of furniture* as compared to *a television is a piece of furniture.* The same occurs with a property statement like *a chair has a seat,* which takes participants less time to verify than *a chair has a back-rest* (Loftus, 1973; Ashcraft, 1978a). Furthermore, gradience effects have been found for items which do not belong to a category. For example, related negatives like *a bat is a bird* and *a whale is a fish* take longer to compare and reject than completely unrelated negatives like *a table is a bird* or *a house is a fish* (Smith, Shoben & Rips, 1974; Glass & Holyoak, 1975). A classical approach like Anderson's (1991a) theory of disjoint boundaries between category memberships cannot explain this gradience in negative items.

All these tasks involve items as members of categories, so assuming that individual members are represented in memory, the gradiences in production frequency data might also be a reflection of that representation, rather than the product of mathematical laws. Also, participants show varying degrees of agreement about an items's membership in a category, with mid-typicality items eliciting less agreement than highly typical or very atypical exemplars (McCloskey & Glucksberg, 1978; Smith, Shoben, & Rips, 1974; Barsalou, 1983, 1987). Thus, production frequency might be taken as reflecting degrees of participant agreement on what the order of exemplars in a category extension should be. If such is the case, then meaningful structure underlies the varying production frequency values of exemplars in a category's extension.

3.04 PROBABILITY OF PRODUCTION OF DIFFERENT EXEMPLARS: PREDICTIONS

One way to distinguish between meaning-based accounts (Smith, Shoben & Rips, 1974; Rosch, 1978) and one which explains gradience as the product of sheer force of numbers (Anderson 1991a), is to look at patterns of participant agreement. All the production frequency data collected in Experiment 1 will be assessed according to the number of different exemplars produced at each of ten levels of production frequency. Patterns of increment or decrement in participant agreement can be examined in the number of different exemplars produced and distributed across ten production frequency levels. For example, level one contains the number of different exemplars with a production frequency of 90-100, which were produced by 91% to 100% of the participants. Level ten represents the number of different exemplars with a production frequency of 1-10, which were produced by 1% - 10% of the participants. The number of different exemplars produced at each level could be reflecting the effects of participant sampling probabilities if one pattern of increment across levels is found; or the influence of participant agreement if a pattern of decrement is found from levels 1 through to 10. The patterns will be described below.

Anderson's (1991a) account sees the order of items in a production frequency task as the product of random effects, with individual items having no predictable results when considered across a large population of subjects. The semantic relatedness account, on the other hand, treats items as meaningful units, where some are more equal than others, making predictions about individual items according to their degree of typicality or their membership. The main reason for the lack of information in Anderson's rational model is that it claims no conceptual representation, it is a simple clustering algorithm. The algorithm groups together objects in a way similar to what people might do, but does not describe any meaningful representation of the objects in memory (Murphy, 1993a).

Thus, the two accounts differ in their use (or non-use) of mental representation of individual exemplars. Because of this difference, the semantic relatedness account predicts a conditional probability between the numbers of different exemplars produced and production frequency levels of items in a category-extension; whilst Anderson's (1991a) account predicts an unconditional sampling probability as the basis for people's production of different exemplars.

Regarding production of same exemplars, Anderson would reason that because all items in a category are theoretically equal members, there is no reason why *the majority* of participants should choose one specific item above the others, except that a participant might use and need some items more frequently than others. Because membership in a category is assumed to be equal for all items, the varying production frequencies for exemplars cannot be explained as being due to one exemplar being a better member than another. Instead, the probability of selection of an exemplar by a participant is governed by individual vagaries of learning experiences across participants, and there is no reason why one person's order of production for the items in a category-extension should closely resemble the order of items produced by another person. The 91-100 level of production frequency signifies that the majority of participants were producing at this level, and simple mathematics would predict that the number of different items produced will increase as the number of people who are doing the producing increases. Consequently, it is this level of production frequency (the majority of participants) which should produce the highest number of different items.

If, on the other hand, the probability of production of the same exemplar is conditional on all participants consulting the same mental representation, then the probability that the same exemplars will be produced will increase with the production frequency level of the exemplar. Assuming that a low production of different exemplars reflects participant agreement about what constitutes an appropriate exemplar, and that people are consulting the same mental representation, then the number of *different* exemplars produced will decrease as production frequency levels increase. In such case, the 91 - 100 level should show the lowest number of different items.

This prediction mirrors a semantic relatedness account, and can also stem from a Roschean approach. Rosch would argue that because exemplars differ in their degree of membership and typicality, then *the majority* of participants should choose one specific item (or a small number of the most typical) above the others (Mervis, Catlin & Rosch, 1976). As a result, the probability of a different item being generated will decrease as a function of the increasing number of people agreeing about its typicality.

Regarding the three category-types and their category-size (that is, the number of different exemplars in each), both Anderson and Rosch would assume that superordinate categories contain a larger number of different exemplars (that is, a larger category-size) than either property or ad hoc types. Anderson (1991a) would predict that when the sample is very large (in memory), then the production of items different from another person's is also high. Participants are likely to disagree *more* about a larger size category-type, because each person has a greater variety of exemplars from which to choose.

Using Rosch's (1978) bases for categorization criteria, people who live in the same culture will tend to agree more about everyday items in a world which they all share. The more the categories in a category-type consist of norms shared by the whole community, the fewer the variety of different items selected, because participants agree more. Property categories are less likely to be represented as such by a large number of people, so a higher number of different items would be generated for them, in comparison to superordinates. Finally, ad hoc classes are by definition new categories so people will produce the greatest number of different items.

In conclusion, Experiment 1 examines two questions:

(a) are there *significant* differences between production frequency levels of the exemplars in a category extension (that is, gradience); and if these are found,

(b) what is the underlying cause of this gradience in the production frequency data? Is it meaningful structure about which most participants will agree, or is the order of the exemplars a product of learning through individual experience? The participant population will be divided into ten levels of production frequency and the data re-assessed accordingly, based upon how many different items people produce at each level.

3.05 METHOD

Participants:

One hundred university students participated in the study, ranging in age from 17 to 60 years.

Stimuli Labels

Each stimulus label belonged to any one of three category-types: natural superordinates, property and ad hoc. Examples of categories which belong to natural superordinates are *Vegetables, Weapons;* examples of property types include *Grating Sounds, Comfortable Things;* and examples of ad hoc types include *Traits which will facilitate friendship, Things to do for weekend entertainment.* For the full list of category-labels and their lists of exemplar production frequencies, see Appendix D.

Materials and Design

Data were collected in booklets, with one category-type per page, consisting of superordinate, property and ad hoc category-types. Each page contained nine category labels and eight blank lines under each of these labels for exemplars to be generated by the subject. The first page consisted of name, address and instructions. The order of the pages was the same for each booklet, so that order of presentation of category-types was invariably natural superordinate, property, ad hoc. Since participants had been asked to produce eight exemplars of each category, with twenty-seven categories in all, each person was producing 216 exemplars.

Task and Procedure

The booklets were handed out to participants, who filled them out at their leisure, and returned them to the experimenter when complete. The task was exemplar generation, in response to a category label, for example *Vegetables.* Participants were asked to "generate eight examples of each category label in the order in which they come to mind", and to write the exemplars down in that order. No time limits were assigned for the task.

3.06 RESULTS

3.06.1 First analysis: production frequency of same exemplars as the dependent variable.

In the first analysis, the dependent variable was the frequency with which a specific exemplar was produced. For each category, across the 100 participants, the most frequently produced exemplars were chosen and listed, with their production frequency values. See Appendix D for these norms. This first analysis was concerned with the production frequency data for the ten most frequently produced exemplars. Each value represented how many people out of 100 had produced a particular item as an example of its category. Invariably, exemplars of a category are not produced with the same frequency. Production frequency was chosen as a measure of gradedness in the category's extension, because it makes fewer assumptions about what specific causes might underlie such gradience. The order in which exemplars were produced by each individual participant was not noted.

Consequently, the items for each category list were ordered according to how frequently each item was produced across the one hundred participants. For example, since the item *chocolate* was mentioned by the highest number of persons (83) for the category *Things not to eat on a diet*, then that item was taken to be the most representative of its category, and first in the order of item dominance. Although other instances of the category were produced first by a number of participants, and *chocolate* was sometimes produced second, third, or fourth by many people, it was the item produced *most frequently* across all participants, and so was taken to be the most typical instance in the category norm.

The first question of interest concerns the relative production frequencies of each exemplar in a category, and whether the category exemplars extensions demonstrate gradience in the frequency with which they are produced. Do all three category-types (superordinate, property and ad hoc) show significant gradience effects between levels? Furthermore, do the most frequently produced exemplars for property and ad hoc categories show production frequency means equal to those of the exemplars of natural superordinate types?

To assess this, a 3 (category-type) x 10 (exemplar-level) analysis of variance was carried out on production frequencies of the ten most frequently produced exemplars in each category. These exemplars, arranged in order of their production frequency values, represented the top ten levels of production frequency: level 1 consists of the most frequently produced of all exemplars and level 10 consists of the tenth most frequently produced exemplars. The exemplar frequencies of items produced at the eleventh, and beyond, levels are not included in this first analysis.

For each of the top ten most frequently produced exemplars in each category, values were calculated that represented the number out of 100 that produced the exemplars. These values were used as the dependent variable, and are presented in Table 1 below.

Table 1. Mean rating scores in the production frequency of exemplars in the top ten range (N = 100).

Exemplar Level*	Superordinates		Properties		Ad Hocs		Means	
One	82.01	*12.56*	65.78	*20.55*	68.89	*14.83*	**72.22**	*17.25*
Two	75.44	*15.07*	52.11	*14.89*	58.56	*10.83*	**62.04**	*16.57*
Three	66.01	*13.43*	46.01	*11.18*	51.78	*8.15*	**54.61**	*13.71*
Four	52.67	*10.47*	42.11	*12.05*	45.01	*7.75*	**46.61**	*10.82*
Five	49.67	*10.71*	37.67	*10.73*	40.44	*7.42*	**42.61**	*10.75*
Six	43.78	*9.91*	32.89	*7.36*	34.44	*6.44*	**37.04**	*9.15*
Seven	39.11	*8.08*	28.33	*5.51*	31.11	*7.49*	**32.85**	*8.27*
Eight	33.67	*6.26*	26.11	*4.88*	28.33	*6.31*	**29.37**	*6.49*
Nine	29.78	*4.79*	23.56	*2.65*	27.01	*5.77*	**26.78**	*5.12*
Ten	22.89	*5.92*	22.22	*2.44*	24.44	*4.42*	**23.18**	*4.42*
Means	**49.51**	*21.18*	**37.68**	*16.81*	**41.01**	*16.28*	**42.73**	

Category-Types

*Exemplar Level = Ten levels represent ten most frequently produced instances of each category (nine categories in each category-type).
Italics = Standard Deviations; **Bold** = Means.

A x B Interaction

Referring to Table 1, the category type variable did not interact with the exemplar-level to a significant degree, with F (18,240) = 1.184, $p > 0.05$. The lack of an interaction means that the three types do not differ from one another in the steepness of their frequency gradients across the ten levels, and suggests that the category-types do not differ in the degree of definedness (or salience) of their category-items.

Variable A: Exemplar-level of production frequency

The difference between the various exemplar-levels was highly significant, at F (9,240) = 72.462, $p < 0.01$. This result indicates the presence of significant gradience effects across the ten exemplar levels, but cannot be taken as indicative of an underlying structure. The items were relegated to these levels as a function of their frequency of mention across 100 participants (not the specific order in which they themselves were produced). Consequently, inferences cannot be drawn as to agreement about *the order* of production of the items, such as which item should be produced first, second, or tenth.

The means for the ten exemplar-levels of production frequency data were subjected to a post hoc analysis to assess whether the increment between levels was significant, using a Newman Keuls test of comparison between means of exemplar-levels of Wr (2,240) = 2.73. In the superordinate types, each exemplar-level incremented significantly to the next, beginning at level 10 and ending at level 1 of production frequency. In the property types, levels 10, 9, 8, and 7 showed no significant increment between levels, although levels 10 and 7 were significantly different (Wr 4,240 = 3.57). From levels 7 to 1, the production frequency data at each level showed significant gradience effects between levels (Wr (2,240) = 2.73). For the ad hoc types, levels 10, 9, and 8 did not show significant gradience effects between levels, though level 8 did increment significantly from level 10 (Wr (3,240) = 3.31), but the pattern changed for levels 8 through to 1, with increments in production frequency being significantly different at each level (Wr (2,240) = 2.73). Considering these post hoc results, all that can be said about this main effect is that some items were produced more often than others, indicating greater agreement on the more frequent items, as they were chosen more often in the context of a particular category.

The standard deviations in Table 1 show a steady decline in variability from exemplar-level 1 to level 10, with decreasing frequency of exemplars produced. This variation in standard deviations could reflect variation across the nine categories used (in each category-type) in terms of the most salient exemplars. For example, the most frequent exemplar (*potatoes)* in *Vegetables* was produced by almost everyone (production frequency = 90); while this was not the case for the most frequent exemplar of *Fish* (*shark* production frequency = 70). So the decline in standard deviations at each level might reflect differences between categories (in the same type) in terms of the number of potential exemplars available to each.

<u>Variable B: Type of category</u>

Collapsing across the ten levels, category-type differences were significant, at F (2,240) = 34.489, $p < 0.01$. The superordinate category-types elicited a higher production frequency of exemplars (49.50), than did property (37.68) or ad hoc (41.00) category types.

A post hoc analysis was carried out on the category-type means to determine the source of variability, and all three category-types were found to be significantly different from one another. Variability in properties was different to that in ad hocs, as were those between ad hocs and superordinates (Wr (2,240) = 2.73); and also between the properties and superordinates (Wr (3,240) = 3.31). The post hoc results signify that participants agreed more about the items produced in superordinate category-types, as the exemplars in these types elicited higher production frequencies.

Participants' production frequencies for exemplars were not the same across the three category-types. This result suggests that all three category-types differ in how well-established their exemplars are in participants' memories. Of the three types, the best-established exemplars are those for superordinate category-types. This inference is borne out by the standard deviations at the first level of production frequency: although the property category-type elicited much lower mean production frequencies (65.78) than the superordinate type (82.01), variability in the property categories (20.55) was much higher than the superordinate type (12.56). For example, it is concluded that *potatoes* is a more widely acceptable norm for the natural superordinate type *Vegetables* (production frequency = 90), than is *blood* for the property type *Red Things* (production frequency = 47), with *blood* having other potential categorizations besides the category of *Red Thing*.

3.06.2 Second analysis: production of different exemplars as the dependent variable.

Gradience effects alone are not a certain indication of an underlying graded *structure* in a category. The second question as set out in section 3.04 asked what underlies the gradience in the production frequencies of the various exemplar-levels: differing degrees of category representativeness; or sampling probabilities in the participant population?

To answer this, it was necessary to re-analyse the data. In the previous analysis, the dependent variable (production frequency) was how many people produced the *same exemplar* in response to a category-label (for example, *Vegetables - potatoes* = 90 subjects). In this next analysis, the dependent variable consists of the number of *different exemplars* being produced. The previous analysis examined only category-exemplars with the ten highest production frequencies. Data for this analysis include all the exemplars produced by all participants, at 216 exemplars per subject. Not all these 216 exemplars need be different, since a person might produce *wine* as an exemplar of *Beverages*, and again as an exemplar of *Things to take on a picnic*.

The production frequency values were divided into ten levels of 1-10, 11-20, up to 91-100. The number of different exemplars in a category which had a production frequency of say, between the 21 - 30 range were then summed and allocated to that level. For example, at the 21-30 production frequency level, participants generated six different exemplars for the category *Fish* (that is, *whiting* = 24, *barramundi* = 28, *perch* = 24, *whale* = 23, *cat fish* = 23, *mackerel* = 22). Then again, there were no exemplars at all of *Fish* generated at the 91-100 level. If the norms in Appendix D are consulted, it will be seen that the highest production frequency for a *Fish* exemplar was *shark*, produced by 61 participants.

The use of this kind of data can be taken as an indirect measure of the extent of participant agreement about which would be the best exemplar for each of the various exemplar-levels of gradience found in production frequencies of the previous analysis. In a perfect world, all participants in the 21-30 range would agree about which instance would be the best example for this level nine, and the same exemplar would be produced by all. Thus, the number of *different* exemplars produced by people is a reflection of their disagreement about which is the most appropriate instance of a category at a certain exemplar-level.

The question being investigated in this analysis concerns the probability of people producing different exemplars: will the number of different exemplars increase in the higher production frequency levels, or decrease? If the former is the case, that would be support for the gradiences found in the previous analysis being a statistical artifact, because in this present analysis, the number of different exemplars produced might be a mathematical function of the number of participants doing the producing. A related question concerns category-types: do superordinate, property and ad hoc types differ as to category-size, estimated by the mean number of different exemplars produced per category-type?

To assess these two questions, a 3 (category-type) x 10 (participants at each production frequency level ANOVA was used to analyze the average number of different exemplars generated per category, at ten levels of increasing participant numbers (ten people per level). The mean rating scores of different exemplars are set out in Table 2.

Table 2 Mean rating scores in the number of different exemplars produced at ten levels of production frequency.

Prod.Freq.	Category-Types							
Level*	Superordinates		Properties		Ad Hocs		Means	
91-100	0.33	*0.71*			0.11	*0.33*	**0.15**	*0.46*
81- 90	0.89	*0.78*	0.44	*0.76*	0.11	*0.33*	**0.49**	*0.70*
71- 80	1.33	*2.24*	0.11	*0.33*	0.33	*0.5*	**0.61**	*1.39*
61- 70	1.22	*1.09*	0.44	*1.01*	0.66	*0.86*	**0.78**	*1.01*
51- 60	1.22	*1.09*	0.65	*0.86*	1.33	*0.71*	**1.07**	*0.91*
41- 50	1.88	*1.45*	1.44	*1.33*	2.01	*1.01*	**1.78**	*1.25*
31- 40	1.77	*1.39*	2.55	*1.66*	2.22	*1.98*	**2.18**	*1.66*
21- 30	4.11	*2.02*	4.55	*1.81*	4.22	*2.04*	**4.31**	*1.90*
11- 20	9.66	*3.81*	13.44	*3.43*	12.01	*4.03*	**11.71**	*3.97*
1 - 10	25.67	*15.58*	47.01	*10.45*	42.66	*8.47*	**38.44**	*14.78*
Means	**4.81**	*8.96*	**7.07**	*14.35*	**6.57**	*12.92*	**6.15**	

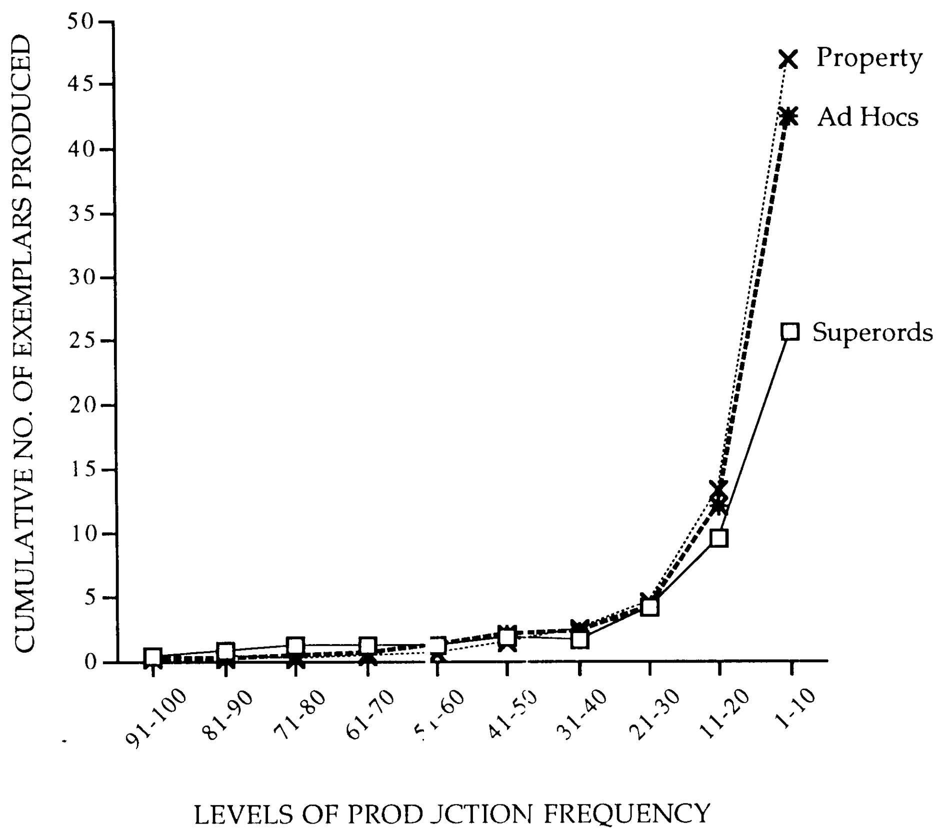
*Prod.Freq. = Production frequency values, with range of 10 at each level.
Italics = Standard Deviations; **Bold** = Means

A x B Interaction

Category-type did interact significantly with the subject-level of different exemplars, with F (18,240) = 7.016, $p < 0.001$. This significant interaction indicates that the underlying basis for gradience effects is structure of some kind, and that it varies across the three category-types. For a visual representation of the number of different exemplars produced at each of the ten levels of production frequency, see Figure 1.

As can be seen in Figure 1, the number of different exemplars *increased* as the production frequency levels decreased. In other words, the number of different exemplars decreased as the number of participants doing the producing at each level increased. In each of the three types, there was a number of levels at which the rate of different exemplar production barely changed, indicating a lack of steepness in gradience for those levels.

Figure 1.



Mean Number of different exemplars
produced at ten levels of production frequencies

Variable A: Category-Type

Referring to Table 2, the superordinate category-types elicited a lower number of different exemplars (4.81) than either property (7.07) or ad hoc (6.57) category-types. Category-type was significant at F (2,240) = 7.513, $p < 0.001$.

Means for the separate category-types showed that property types had the largest category size, then ad hoc types, and then superordinate types. Post hoc tests reveal that the superordinates were significantly different to the property types (Wr (3,240) = 1.36) and to the ad hoc types (Wr (2,240) = 1.14). The property and ad hoc types means did not differ.

The results of the first analysis, set out in Table 1, suggested that the superordinate categories might be better established in memory than the other two, and this is confirmed by the present analysis. The natural superordinate types, in spite of having the highest production frequencies, were found to produce the lowest number of different exemplars. If a *shared* mental representation of *norms* was the cause for the results of the previous analysis, then in the present analysis, the superordinate category-types should show a smaller category-size than the other two category-types, because subjects would disagree less about which natural superordinate exemplars are the "representative" ones and so choose those.

Variable B: Number of different exemplars at each Production Frequency level

The difference between production frequency levels for producing the different items was significant, at F (9,240) = 225.883, $p < 0.001$. Collapsing across category-types, there was a systematic decrease in subject *dis*agreement, with the number of different items produced steadily *decreasing* as the number of participants doing the producing increased (see Figure 1).

The Newman Keuls post hoc analyses showed that significant decrement in the number of different items produced stopped at the range of 31-40, and neither increased or decreased up to the range of 91 - 100. Significant increases in disagreement occurred amongst participants from range 31-40 to range 1 - 10, using Wr (2,240) = 1.14. This pattern was exhibited in all three category-types. The results of this statistical analysis strongly indicate that the gradience effects found in the production frequency task of the foregoing analysis should now be interpreted as being based upon an internal gradient structure in each category's extension.

Overall, the main outcome of this analysis is that production of different exemplars did *not* increase as a monotonic function of the increasing number of participants doing the generating, which would be the case if gradience was a statistical artifact of the number of participants in each level.

The method of estimating participant agreement according to the number of different exemplars produced was taken from Barsalou (1983), who measured graded structure in both natural categories and ad hoc categories by participants' performance on production frequency and typicality rating tasks. Both these forms of graded structure are well correlated in natural categories (Mervis, Catlin & Rosch, 1976). Salient typicality gradients have always been found in these categories, but Barsalou (1983) showed that ad hoc categories also had salient *graded* structure. He estimated salience of graded structure by participant agreement, reasoning that if ad hoc categories do not possess salient gradient structures, then participants should show no agreement about items at each level of gradience. His results were similar to the pattern of increment found here, where there is a relationship between incrementing agreement and increasing numbers of participants. The data from Experiment 1 for the property and ad hoc types show evidence of an exemplar-order about which most people can agree, at least for the first 6 to 8 levels.

3.07 DISCUSSION: CATEGORY-TYPE DIFFERENCES AND REPRESENTATIONS

The data analyses do not support the hypothesis that the gradience effects found in most category-types might be the result of sampling probabilities and category-size. The results do support the hypothesis that the gradience effects found in categories have an underlying structure. The majority of people agreed about which exemplars of a category are the most appropriate ones, as shown by the decreased number of different exemplars at the higher levels of production frequency. It is likely this agreement was due to their consulting a common conceptual representation of items in the category-extension. The results of the second analysis also imply that the nature of this structure might vary with each category-type.

3.07.1 Category-type differences

Differences in degree of established category

The main outcome of the first analysis was a demonstration that all three category types differed significantly in their rates of mean production frequency which they elicited from participants. The natural superordinate labels elicited exemplars with higher production frequencies than either property or ad hoc category-types. This implies that superordinate exemplars are better established as category units, with ad hoc categories second, and property categories last. Natural categories have been termed "common" or "everyday" categories, and their everyday occurrence might explain why their items are so well-established in most participants' memories, in comparison to ad hoc or property types (Rosch, 1973; Barsalou, 1983; Barsalou & Ross, 1986).

The production frequency means obtained in this first analysis for the superordinate and property category types might be explained by a prototype model of category structure. Rosch (1978) described natural categories as reflecting *perceived* world structure, and since all people share the same fuzzy perceptions of world structure and its natural categories, they are more likely to agree about their exemplars. Subjects are more likely, therefore, to produce the same exemplars as other subjects in these categories because they share the same representation of the category concept. From this, it might be expected that exemplars in superordinate category-types would have the highest production frequencies, and property exemplars would have lower production frequencies because they are less well-established as representations. The absence of an interaction in this analysis indicates that the decline in production frequency gradient was roughly the same for all top ten exemplars in all categories.

One unexpected result is the production frequency mean for instances of ad hoc types, which was significantly higher than that for property types. Property category-types (for example, *Loud Sounds* or *Poisonous Things*) would have some featural information stored in semantic memory about their referent objects, and should elicit higher means for their exemplars than ad hoc types. Since the participants presumably compiled lists of ad hoc exemplars for the first time when presented with the category-label in the booklet, the exemplars of the ad hocs could not be represented as a coherent category structure. So it cannot be concluded that the ad hoc category exemplars had a higher production frequency rate because they themselves are better established *as a concept* than the property types. For example, people might represent a goal-driven schema for information on "Diets, and how to lose weight", which can be used to construct a list when required, in response to *Things not to eat on a diet.* Thus, property types appear to be less-established in memory than ad hocs, because participants might not have background information or schemata upon which to draw when generating property type exemplars.

Differences in category-size

Concerning category-size (mean number of different exemplars), Anderson (1991a) uses category-size as a base rate, and predicts that the larger the sample available in memory in a category-type, the less probable that a person would produce the same exemplars as another subject. He therefore predicts that superordinates will have the largest mean for different exemplars produced across the three types.

The second analysis of the data from Experiment 1 showed this not to be the case, with results supporting a Roschean hypothesis. This states that the more everyday the items in a category, the better established they are likely to be as norms shared by the whole community. Consequently, although superordinate's category-size might be larger, most people would agree about which were the better exemplars and tend to generate those, resulting in a smaller mean for category-size in superordinate types.

3.07.2 Participants share a common representation

The results of the second analysis can be used to draw inferences about what participants were doing in the exemplar-generation task. In the first analysis, the production frequency values also represented participant percentages, that is, how many people (out of 100) produced a specific exemplar. In the second analysis, the number of *different* exemplars which fell within the range of production frequencies (say, 21 to 30) were calculated, and this number of different exemplars was taken as reflecting participant disagreement about which was the best example (for that range of production frequency).

It can be inferred from the outcome of the second analysis that the differences in exemplar levels of production frequency (found in the first analysis), were based upon an underlying structure in the category extension. In order to reach the opposite conclusion, that sheer force of numbers was generating the gradience effects in production frequency, a consistent increase of different exemplars at each level should have been found in the second analysis (see Table 2 in Results section 3.05.2). Instead, the number of different exemplars produced, in natural superordinates, barely changed until the eighth production frequency level (21-30) was reached. From this eighth level onwards, the number of different exemplars increased sharply. In the property types, the number of different exemplars produced did not markedly increase until the 6th level (41-40); and in ad hoc types a sharp increase began at the 5th level (51-60). In short, the number of different exemplars became fewer as the number of participants increased, suggesting that subjects were consulting a representation shared by most of them (Barsalou, 1983), especially in the case of natural superordinate types.

That the three category-types differed significantly in their underlying structures was confirmed by the presence of an interaction between the production frequency level and the category-type (see Figure 1). Consequently, the "steepness" of the gradience differed across the three category-types, and occurred most sharply from levels 8 to 10, with property and ad hoc types beginning earlier and producing much higher numbers of different exemplars at those levels. From levels 8 to 10, it would seem that the gradience in the number of different exemplars became much steeper, suggesting that representations were less shared across participants, at least for the property or ad hoc types.

The empirical evidence here contradicts Shipley's (1993) claims that ad hoc categories should be described as "classes" rather than "categories". According to Shipley's (1993) distinction between "classes" and "categories", ad hocs are the former, not the latter. She gives a number of reasons for not defining them as categories: the labels of ad hoc categories would never be used for the identification of an isolated object; their members are not thought to share a deep resemblance; and the physical features of an instance are not naturally attributed to other members. Barsalou (1987) grants that these ad hoc lists may be temporary constructs in working memory created once to support decision-making related to some goal-directed behaviour. Yet, he points out, their items do exhibit coherence as a comprehensible category, even though they might not be physically similar. There is no relational or physical similarity shared by the items in a list like *children, stereo, wallet* and *car.* When a label like *Things to save from a burning home* is applied, however, the items become comprehensible as a coherent concept. Barsalou (1983) claims that all the instances are sharing a common dimension known as the "ideal goal". In *Things* *to save from a burning home* this "ideal goal" would plausibly be *to eliminate loss* (whether the loss be sentimental intrinsic, familial or financial).

To summarise the conclusions drawn from Experiment 1, the results indicate the presence of meaningful structure underlying the production frequency distributions in each of the three category-types. The main point which can be inferred is that people were consulting a shared representation of the category whilst carrying out the task of exemplar production (Barsalou, 1983). The nature of the data, however, does not allow further inferences to be drawn about what kind of representation underlies the production frequencies of the exemplars. The significant interaction of the second analysis would indicate that this underlying representation differed across the category-types. Natural superordinates produced a more gradual gradedness in subject disagreement than the property and ad hoc types. Again, however, valid inferences cannot be made about the kinds of representations (for example, as structural prototypes, featural definitions, or dimensional schemata) which might differentiate the category-types one from another.

Concerning the individual category-types, the first analysis suggested that natural superordinates have exemplars which are more salient and production frequency gradiences which are steeper than the other two types. Their category-size would seem to be smaller, but this result is reflecting higher levels of agreement amongst participants, as to which instances make better examples of a category, rather than the actual size of the category. The property types had more salient exemplars than the ad hocs (as shown in analysis one); but their structure has roughly the same gradience or steepness. Their category-size was the same in that people seem to agree to the same extent (as shown in analysis two).

3.08 IMPLICATIONS

3.08.1 Category represented as a structure or as content

As Figure 1 shows, the gradience of the participant agreement in analysis two did not change sharply from one level to the next. One possible implication of this result is that the production frequencies of analysis one do not reflect *graded* structure at all. Instead, they possibly reflect category content (i.e., normative knowledge). The most plausible interpretation of the second analysis results remains that participants were consulting a common representation, but considering the lack of steepness in the incrementation in number of different exemplars at each level, it is possible that such a representation was based on normative knowledge about the items. If it had been a *structural* representation of typicality, the gradience would have been steeper, with significant differences in agreement amongst participants occurring at each production frequency level.

This possible explanation for the lack of gradience in the participant agreement of the second analysis was suggested by a paper on social epistemology by Freyd (1983), who claims that people do not represent normative structure. Rather, the significant differences between exemplar-levels in people's production frequencies might simply be a structure which has "emerged" as a result of people's use of stereotypes. The results might be reflecting their use of a common structure, but not necessarily one which is *mentally* represented. This possibility was not considered by Anderson (Anderson, 1991b; Corter, 1991).

Freyd's (1983) argument is that many observed structures such as categories are psychologically real, but they do not necessarily need to be internally representable. Rather, the structures are reflecting shareability constraints on human knowledge, with the structures emerging from the problem of sharing knowledge with other minds (not necessarily imposed by the individual mind). Thus, the need to communicate, as well as the need to understand, imposes shareability constraints.

The main form the constraint would take would be to keep the knowledge structure simple. One strategy which is used in categorization is to note the presence or lack of features which might define or characterise the category. A simple category structure would be one with few feature or attribute dimensions, with a small number of values on each dimension. Although many dimensions might be available for categorizing real objects or abstract ideas, people would tend toward isolating a few dimensions known to all. They could apply these dimensions to a number of knowledge domains, to ease the problem of agreeing on, for example, which categories certain objects belong to, or which objects are better exemplars of it. In this way, ease of shareability would begin to shape the knowledge structure. Such an "emergent structure" might be represented originally as a dimensional schema, as suggested by Barsalou (1983) for ad hoc categories.

Shareability constraints might be expressed in the form of people's use of analogies in the explanation of new concepts. For example, if someone describes Walter as looking like a giraffe, the listeners are not likely to think that Walter has the eating behaviour or the colouring of a giraffe, but they are most likely to refer to the giraffe's salient feature, which is its long neck, thereby assuming that Walter also has a more-than-average long neck (Freyd, 1983). As a result, because people share the background knowledge that giraffes are animals with exceedingly long necks, it is known that this is the feature to focus upon (rather than its eating behaviour or colour).

The kind of emergent, normative structures described by Freyd (1983) are reminiscent of Putnam's stereotypes, which were described as concepts shared and developed by lay people from the scientific knowledge of experts. Examples include the scientific definitions of the true essence of gold, or water, or lemons, which lay persons might understand only vaguely or not at all (Putnam, 1975a, 1975b). Freyd (1983) says normative knowledge reflects social epistemology, and the "emergence of structures" through people's common use of them. She repeatedly notes that she is not specifying anything about constraints on internal representation *within* the individual human. In her theory, knowledge has an individually-based representation. One way to reconcile her theory (and Putnam's) of a social epistemology with the results of Experiment 1 would be to look at the theories of dual representation in categories, which have surfaced in the research literature since Murphy and Medin's paper in 1985. If the theory-based models of representation are supported, the results might show that people have dual representations based upon different kinds of knowledge.

The notion of shareability of structures is necessary to explain how people manage to communicate, but there is a paradox in that each person's knowledge and experience is unique to that individual (at least theoretically). Thus, a person must have a "common understanding" of a prototype or stereotype in order to communicate, but surely also has a knowledge that is unique to self, the latter probably derived from emotions, needs, and individual intentions. According to Freyd (1983), it is this idiosyncratic knowledge which is represented mentally in a category.

One drawback in the use of normative structures is the loss of distinctive features of the information, resulting in distortion effects. As the size of the community of knowledge sharers increases, the distorting effects should increase, in the way often observed in the spread of rumours. This is not a new idea, as Bartlett's stories demonstrated, when he performed a number of "serial reproduction" experiments as far back as 1932. Each participant was asked to repeat a story which she or he had just heard, to the next participant. The stories were distorted, as happens with rumours, with information being re-organized to fit the listeners' schemata. Bartlett (1932) pointed out that schemata are conventions and quite possibly arbitrary conventions, but once in place they impose structure upon the information about an item which is being incorporated.

3.08.2 Comparison of idiosyncratic and normative stimuli in Experiment 2.

To Bartlett's interpretation, Freyd (1983) adds the prediction that, not only would the information be re-organized, it would be distorted in such a way as to make the structure a more simple dimensional representation of the stimuli. One way in which Experiment 2 was designed to investigate the possibility of a distorted and impoverished representation or categorization process was to compare the effects upon task performance of stimulus-items which are either idiosyncratic or normative. They might affect category behaviour in different ways, such as making more errors in the processing of normative items.

What sort of experiential knowledge best represents categories? Does each person organize his or her category knowledge in an idiosyncratic way, or do we all think alike in a normative way? Anderson (1990) and Rosch et al. (1976) both described a rationalist mind which, to varying degrees, was influenced by the outside environment, and in that sense might be said to have experiential knowledge. Such knowledge, in the case of Anderson's theory (1990), is *objective* because it is the same experience shared by all people in the one "true" world. In the case of Rosch et al. (1976) such knowledge is *normative* because it is the same experience shared by all people in the same culture.

Alternatively, some experiential knowledge is both *idiosyncratic* and *subjective*, in that its categorical organization is based upon personal interpretations of direct experience, rather than cultural constraints or environmental determinants (Barsalou, 1992). It is doubtful that idiosyncratic knowledge is stored separately from normative information in the individual's mind, yet most early research in the field made use only of artificially constructed categories, typicality norms and production frequency norms as stimuli. By definition, all these are very objective and abstract.

Normative data is generated across a wide range of subjects, and often used as stimulus materials in further experiments. Rosch's (1973) experiment not only provided evidence for exemplars' degrees of typicality in natural categories, but ensured their use in many experiments since, as the fixed degree of typicality which a word possesses. The exemplar *robin* has repeatedly been cited in the literature as the most typical member of the category *Bird*. Yet as the exemplar list for *Bird* shows, in this experiment, *robin* did not make the first ten exemplars (see Appendix D). This reflects the differing experiences with birds for persons in this study and those (North American or English) participants in other studies. Another example is the Battig and Montague (1969) production frequency norms, which were generated by giving 442 subjects a category name or description, and asking them to write down as many category members as they could think of in a thirty-second period. The items which were most frequently produced were those considered to be most typical of their category. In short, the essence of normative data is a reflection of what is most people's "accepted norm" for typicality of items (or their production frequency) across a population. Its advantages are that it provides a stable and invariant structure for a normative category understood by all.

The problem with using normative stimuli in categorization tasks and other measures of internal structure is that the individual's idiosyncratic knowledge is lost, and this *subjective* aspect of the information might be what is most important for the coherence of categories. Direct experiential or autobiographical knowledge might influence the processing and organization of categories (for example, ad hoc categories) in ways that are different to objective and/or normative knowledge (Conway, 1990; Kahnemann & Miller, 1986). Subjective experience involving emotions, needs, intentions, may be the "glue" which coheres the items of a category closely, and it might allow for a more distinctive account to be made of an individual's use of categories (Lakoff, 1987).

In conclusion, Experiment 1 has shown that category-members are mentally represented, but the data does not make clear what form this representation might take. It is possible, taking into consideration the lack of gradience in participant's agreement about good examples, at least in the first 6 to 8 levels, that participants' representations of a category might be more in the nature of a knowledge structure than a structure of physical appearance. If the former is the case, then individual experience might be of relevance in category representations. This issue is examined in the following chapter.